Combined untargeted and targeted fingerprinting by comprehensive two-dimensional gas chromatography: revealing fructose-induced changes in mice urinary metabolic signatures

Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography: Revealing Fructose-Induced Changes in Mice Urinary Metabolic Signatures

Davide Bressanello[1$], Erica Liberto[1$], Massimo Collino[1], Fausto Chiazza[1], Raffaella Mastrocola[2], Stephen E. Reichenbach[3], Carlo Bicchi[1] and Chiara Cordero[1*]

Authors' affiliation:

[1]*Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino,*
Via Pietro Giuria 9, I-10125 Torino, Italy

[2] *Department of Clinical and Biological Sciences, Università degli Studi di Torino,*
Corso Raffaello 28, I-10125 Torino, Italy

[3]*Computer Science and Engineering Department, University of Nebraska*
1400 R Street, Lincoln, NE 68588-0115, USA

§ Davide Bressanello and Erica Liberto, listed in alphabetical order, equally contributed to this work

* Address for correspondence:
Dr. Chiara Cordero - Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy – e-mail: chiara.cordero@unito.it ; phone: +39 011 6707662; fax: +39 011 2367662

Abstract

This study exploits the information potential of comprehensive two-dimensional gas chromatography configured with a parallel dual secondary column-dual detection by mass spectrometry and flame ionization (GC×2GC-MS/FID) to study changes in urinary metabolic signatures of mice subjected to high-fructose diets. Samples are taken from mice fed with normal or fructose enriched diets provided either in aqueous solution or in solid form and analyzed at three stages of the dietary intervention (1, 6, and 12 weeks).

Automated *U*ntargeted and *T*argeted *fingerprinting* for 2D data elaboration is adopted for the most inclusive data mining of GC×GC patterns. The *UT fingerprinting* strategy performs a fully automated peak-region features fingerprinting and combines results from pre-targeted compounds and unknowns across the sample-set. The most informative metabolites, with statistically relevant differences between sample groups, are obtained by unsupervised multivariate analysis (MVA) and cross-validated by multi-factor analysis (MFA) with external standard quantitation by GC-MS.

Results indicate coherent clustering of mice urine signatures according to dietary manipulation. Notably, the metabolite fingerprints of mice fed with liquid fructose exhibited greater derangement in fructose, glucose, citric, pyruvic, malic, malonic, gluconic, cis-aconitic, succinic and 2-keto glutaric acids, glycine acyl derivatives (N-carboxy glycine, N-butyrylglycine, N-isovaleroylglycine, N-phenylacetylglycine), and hippuric acid. Untargeted fingerprinting indicates some analytes which were not *a priori* pre-targeted which provide additional insights: N-acetyl glucosamine, N-acetyl glutamine, malonyl glycine, methyl malonyl glycine, and glutaric acid. Visual features fingerprinting is used to track individual variations during experiments, thereby extending the panorama of possible data elaboration tools.

Key-words: comprehensive two-dimensional gas chromatography-mass spectrometry; parallel dual secondary column-dual detection; fructose-induced metabolic derangements; urine metabolic profiling; Untargeted and Targeted fingerprinting

Introduction

The dietary intake of fructose has markedly and rapidly increased over recent decades due to consumers' preferences for sugar-sweetened beverages and sugar-rich processed foods [1]. This trend has been associated with increased prevalence of metabolic diseases, such as dyslipidemia, insulin resistance, and nonalcoholic fatty liver diseases (NAFLD) [2, 3]. All these phenomena are related to the high lipogenic power of fructose [4] that promotes lipid deposition in ectopic tissues, such as liver and skeletal muscle, even without impacting body weight and adipose tissue mass [5, 6].

Recent findings have demonstrated that liquid and solid high-sugar diets differentially modulate feeding behavior, intestinal sugar transporters, and hormone expression [7]. This experimental evidence suggests that not only the type (e.g. fructose vs. glucose), but also the form (liquid vs. solid) of sugars may significantly affect the development of obesity, insulin resistance, and/or fatty liver disease. Despite the paucity of experimental data, this topic deserves better elucidation as sweeteners, including fructose, can be ingested in both liquid and solid formulations.

To date, the specific effects of intake of different forms of fructose, liquid or solid, on intestine integrity and microbiota, and hepatic outcomes, have not been investigated. Hence, this study was designed to elucidate the toxicological profile of high fructose intake in the intestine and liver, to determine the differential impact due to the sugar administration forms [8] and to evaluate its differential impact on urine metabolic signatures. C57 male mice were fed with either a standard diet plus water (SD), a standard diet plus 60% fructose syrup (L-Fr), or a 60% solid fructose plus water (S-Fr), for 12 weeks. Within this context of investigation, urinary metabolic signatures could provide some insights on the mechanisms underlying the toxicological effects revealed by physio-pathological assessments, with further information on the kinetics of the metabolic derangements.

To qualitatively and quantitatively track changes in primary metabolites in the mice urine [9], two-dimensional comprehensive gas chromatography with mass spectrometry (GC×GC-MS) represents one of the most advanced and informative hyphenated GC platforms currently available. Thanks to its separation power, sensitivity, and structured bi-dimensional (2D) patterns for chemically related compounds, detailed profiles and fingerprints of complex biological samples can be comprehensively evaluated [10, 11].

For these analyses, a GC×GC system with parallel dual secondary column-dual detection configuration (GC×2GC-MS/FID) was adopted to maximize the overall separation power and limit second-dimension (^{2}D) column overloading. This configuration improves quantitative reliability and response linearity over a wider concentration range [10, 12]. It operates at close-to-optimal ^{2}D linear velocities in both chromatographic dimensions and doubles the secondary-column loading capacity, with positive effects on overall system orthogonality, efficiency, and selectivity [13–17]. In addition, advanced pattern recognition methodologies that

combine untargeted and targeted investigation for the most inclusive fingerprinting (i.e. UT fingerprinting) [18] by GC×2GC-MS/FID are employed to exploit the information potential encrypted in mice urinary signatures, while also providing convenient tools for visual feature inspection to follow signature changes over time.

Materials and Methods

Reference compounds and solvents

All chemicals were from Sigma-Aldrich (Milan, Italy):

a) pure standards of *n*-alkanes (from *n*-C9 to *n*-C25) for system evaluation and Linear Retention Index (I^T_S) determination;

b) pure standards for quantitative determinations and/or identity confirmation of pyruvic acid, lactic acid, malonic acid, succinic acid, malic acid, 2-ketoglutaric acid, 3-hydroxybutirric acid, fumaric acid, 2-keto-3-metilvaleric acid, aspartic acid, hippuric acid, citric acid, uric acid, L-alanine, L-valine, L-leucine, L-proline, glycine, L-threonine, L-tyrosine, L-phenylalanine, L-isoleucine, L-methionine, L-cysteine, L-ornithine, L-triptophan, xylitol, ribitol, fructose, galactose, glucose, mannitol, myo-inositol, glycerol, palmitic acid, stearic acid, creatinine and the internal standards (ISTDs) 4-fluorophenylalanine (QC for derivatization), and 1,4-dibromobenzene (QC for GC normalization);

c) derivatization reagents O-methylhydroxylamine hydrochloride (MOX) and N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA);

d) HPLC-grade solvents: methanol, pyridine, n-hexane, dichloromethane and toluene.

Samples

Four-week-old male C57BL/6J mice (n=18) (Harlan-Italy; Udine, Italy) were housed in a controlled environment at 25±2 °C. All the animals were fed with a normal pellet diet for 1 week (T1) prior to the experimentation. Mice were divided into three groups: a group fed a standard diet and drinking tap water (CS group, n = 6), a group fed a standard diet and drinking a 60% fructose syrup (FL group, n = 6), and group fed a 60% fructose solid diet and drinking tap water (FS group, n = 6), for twelve weeks. Dietary compositions (Sniff Spezialdiäten GmbH, Soest, Germany) are detailed in **Table 1**. All groups received drink and food *ad libitum*. Animals were cared in compliance with the European Council directives (No. 86/609/EEC) and with the Principles of Laboratory Animal Care (NIH No. 85–23, revised 1985). The experimental protocol was approved by the Turin University Ethics Committee (permit number: D.M. 94/2012-B). Urine samples were collected at 1 week (Basal) and after 6, and 12 weeks (T6 and T12) of dietary intervention, then immediately quenched on liquid nitrogen and stored at -80°C until derivatization/analysis. For urine collection, conscious mice were individually placed in metabolic cages with free access to water for 16 hours.

Urine samples were submitted to a standard derivatization protocol [19] consisting of the following steps: 80 µL of urine and a suitable volume of ISTD (4-fluorophenylalanine solution at 10 g/L) were carefully mixed (Whirlimixer vortex, Fisher Scientific, Loughborough, Leicestershire, UK). Then, the solution was dried under a gentle stream of nitrogen before the addition of 60 µL of MOX. The resulting solution was incubated for 2 hours at 60°C. Next, 60 µL of MSTFA were added and the mixture was incubated at 60 °C for one hour. The resulting sample solution was spiked with 1,4-dibromobenzene at 50 mg/L final concentration and diluted in toluene at a final volume of 200 µL. Analyses were immediately run by duplicate injections for each sample; samples were randomized within a 24 h time frame after derivatization.

GC×2GC-MS/FID instrument set-up

GC×GC analyses were run with the following system configuration: a MPS-2 multipurpose auto-sampler (Gerstel GmbH, Germany) was integrated with an Agilent 6890 GC unit coupled to an Agilent 5975C MS detector (Agilent, Little Falls, DE, USA) operating in EI mode at 70 eV. The GC transfer line was set at 300°C. An *Auto Tune* option was used and the scan range was set to *m/z* 50-450 with a scanning rate of 12,500 amu/s to obtain a spectra generation frequency of 20 Hz. The Flame Ionization Detector (FID) was operated as follows: base temperature 300°C, H_2 flow 40 mL/min, air flow 350 mL/min, make-up (N_2) 20 mL/min, and sampling frequency 200 Hz.

The column set consisted of a first-dimension (^{1}D) column of 30 m × 0.25 mm d_c × 0.25 µm d_f SE52 (95% polydimethylsiloxane, 5% phenyl) connected to two second-dimension (^{2}D) columns of equivalent length of 1.4 m × 0.1 mm d_c × 0.10 µm d_f OV1701 (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl). Connections between the primary and the two secondary columns were by a SilFlow™ GC 3 Port Splitter (SGE Ringwood, Victoria, Australia).The secondary column toward the MS detector was connected to a Quick Swap unit (G3185, Agilent, Little Falls, DE, USA) and to an auxiliary electronic pressure controller (EPC) consisting of a one channel Pneumatics Control Module (G2317A, Agilent, Little Falls, DE, USA). The restrictor capillary in the GC-MS transfer line was of 0.17 m x 0.1 mm d_c. A schematic picture of the system configuration is provided as supplementary information (**Supplementary Figure 1 - SF1**). All columns and capillaries were from Mega (Legnano, Milan, Italy). Helium carrier gas was delivered at constant flow with initial head pressure p_i 255.0 KPa and the auxiliary gas (He) for MS outlet pressure correction was delivered at 39.9 KPa (relative) [13, 14, 20]. The estimated split ratio between MS/FID was 50:50.

Injections for the analysis of both urine samples and *n*-alkanes for Linear Retention Indices determination I^T_S, were by a MPS-2 sampler (Gerstel GmbH, Germany) under the following conditions: split/splitless injector, split mode, split ratio 1/20, injector temperature 290°C, and injection volume 2µL. The oven temperature programme was 60°C (1 min) to 300°C (10 min) at 4.0°C/min.

The system was equipped with a two-stage KT 2004 loop-type thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen and controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). The hot jet pulse time was set at 350 ms, modulation time was 5 s, and the cold-jet total flow was progressively reduced with a linear function from 30% of Mass Flow Controller (MFC) at initial conditions to 5% at the end of the run. Loop dimensions were chosen on the basis of the expected carrier linear velocities to ensure that two-stage band focusing and release were performed for each modulation stage. The first 0.6 m of the ^{2}Ds were wrapped in the metal slit of the loop-type modulator.

GC-MS instrument set-up

GC-MS analyses were run with the following system configuration: an MPS-2 multipurpose auto-sampler (Gerstel GmbH, Germany) was integrated with an Agilent 6890 GC unit coupled to an Agilent 5975C MS detector (Agilent, Little Falls, DE, USA) operating in EI mode at 70 eV. The GC transfer line was set at 300°C. An *Auto Tune* option was used and the scan range was set to *m/z* 35-550 with a scanning rate of 2,500 amu/s. The column consisted of 30 m × 0.25 mm d_c × 0.25 µm d_f SE52 (95% polydimethylsiloxane, 5% phenyl) from Mega (Legnano, Milan, Italy). The carrier gas was helium delivered at constant flow (1.0 mL/min) with initial head pressure p_i 255.0 KPa.

Injections for the analysis of both urine samples, standard calibration solutions and *n*-alkanes for Linear Retention Indices determination I^T_S, were by a MPS-2 sampler (Gerstel GmbH, Germany) under the following conditions: split/splitless injector, split mode, split ratio 1/20, injector temperature 290°C, and injection volume 2µL. The oven temperature programme was 80°C (2 min) to 140°C at 10.0°C/min then to 240°C at 4°C/min up to 290°C (5 min) at 10°C/min.

Raw data acquisition and GC×GC data handling

Data were acquired by Agilent MSD ChemStation ver E.02.01.00 and processed using GC Image GC×GC Software version 2.6 (GC Image, LLC Lincoln NE, USA). Statistical analysis was performed by XLSTAT (Addinsoft, New York, NY USA) and heat map visualization by GENE-E v 3.0.77 (Broad Institute, Inc. Cambridge, MA, USA).

"UT fingerprinting" work-flow

The 2D data elaboration work-flow is illustrated in **Figure 1**. It was optimized and validated in a previous study dealing with the fingerprinting of complex volatile fractions from Extra Virgin Olive oil samples [18]. It combines untargeted and targeted pattern recognition approaches for the most inclusive fingerprinting. In the present study, FID data, although acquired in parallel with MS detection, were not considered because several co-elutions occurred on informative targeted peaks.

2D-data processing used the *template matching* strategy developed by Reichenbach et al. [21]. The approach uses metadata collected from 2D peak patterns (retention times, MS fragmentation patterns, and detector responses) and establishes reliable correspondences between the same chemical entities across multiple chromatograms. The output is a data matrix of aligned 2D peaks and/or peak-regions and related metadata (^{1}D and ^{2}D retention times, compound names for target analytes, fragmentation pattern, single ions or total ions response) available for comparative purposes and further processing [13, 14, 22].

Targeted analysis (Step 1 - **Figure 1**) focused on about 53 compounds reliably identified by matching their EI-MS fragmentation pattern (NIST MS Search algorithm, ver 2.0, National Institute of Standards and Technology, Gaithersburg, MD, USA, with Direct Matching threshold 900 and Reverse Matching threshold 950) with those collected in commercial (NIST2014 and Wiley 7n) and in-house databases. As a further parameter to support identification, Linear Retention Indices (I^T_S) were considered and experimental values were compared with tabulated reference indices. Analyte retention indices, both experimental I^T_S and reference, are listed in **Table 2** together with their retention times (^1t_R min and ^2t_R sec) and relative standard deviations (RSD% over 108 runs).

Untargeted analysis (Step 2 - **Figure 1**) was based on peak-regions features [23, 24] and was performed automatically by GC Image Investigator™ R2.6 (GC-Image LLC, Lincoln NE, USA). The untargeted analysis included all peak-regions above a fixed peak response threshold (5,000 counts for the MS detection channel) together with targeted peaks from Step 1. This approach [11, 14, 18, 23–25] (see also *Section Targeted quantitative fingerprinting*) aligned the 108 chromatograms (6 mice × 3 diet groups × 3 time points × 2 analytical replicates) using a set of registration peaks to produce a composite chromatogram from which feature peak-regions were extracted. Then, the template with these peak-regions and target peaks was aligned with each of the 108 chromatograms for quantitative evaluation. The resulting data matrix was 108 × 380 (samples × reliable peak-regions). Response data were normalized twice: a) Normalized 2D-Peak Volumes were firstly divided by 1,4-dibromobenzene (ISTD) response and then b) over the total image percent response. Aligned 2D peak-regions normalized responses were used for multivariate analysis (MVA) after standardization and results cross-compared to those from target peaks distributions. (See *Section Targeted quantitative fingerprinting* for the results discussion.)

Significance tests were performed on targeted peaks to evaluate the informative role of single analytes in describing the phenomenon under study. The Kruskall-Wallis test followed by Dunn's evaluation was performed on analytes normalized response and considering as sample classes the three diet groups. Results are reported in Table 2 (alpha =0.05).

Visual features fingerprinting, with pair-wise image comparison, also was performed (Step 3 of **Figure 1**) with "colorized fuzzy ratio" rendering mode [26]. The algorithm computes the difference at each data

datapoint (i.e., the output of the detector at a point in time) between pairs of TIC chromatograms. These differences are mapped into Hue-Intensity-Saturation (HIS) color space to create an image for visualizing the relative differences between image pairs in the retention-times plane [26]. A detailed description is provided in *Section "Urine fingerprint evolution by visual features analysis".*

Quantitative profiling by GC-MS: method performance verification

Validation of the GC-MS quantitative method followed a previous protocol [25] on a three-weeks-over-two-months period, characterizing the following parameters: precision, linearity, accuracy, and Limit of Quantitation (LOQ) [25, 27, 28]. Precision data (intra and inter-week precision on retention times and 2D Peak Volumes on analytes target ions - *Ti*) were evaluated by replicating analyses during the entire period. Linearity was assessed by linear regression analyses within the working range, over at least six different concentration levels. Experimental results on linearity, precision (at 10 mg/L), and Limit of Quantification are reported as supplementary information in **Supplementary Table 1** (including analytes target and qualifier ions, calibration ranges, regression curve equations, determination coefficients R^2, limit of quantitation LOQ, and precision on T*i* peak areas at 10 mg/L calibration level with 9 replicates).

The LOQ was experimentally determined by analyzing decreasing concentrations of standard calibrating solutions. Each sample was analyzed in triplicate, and the LOQ was the lowest concentration for which instrumental response (Normalized Peak Area based on *Ti* response) reported an RSD% below 15 %, across replicates [29].

Calibration solutions for quantitative determination were prepared by mixing single-component Standard Mother Solutions at 10 g/L in suitable solvents (acetone or toluene) and adjusting the final volume up to the required concentration. Each solution was then submitted to derivatization (*Section Samples*) and directly analyzed. Calibration levels investigated were: 100 mg/L, 40 mg/L, 30 mg/L, 15 mg/L, 10 mg/L, 5 mg/L, 1 mg/L, 0.5 mg/L, and 0.1 mg/L. 4-fluorophenylalanine, i.e. the Internal Standard for derivatization quality control was at 10 mg/L.

Results and discussion

The goal of the study was to evaluate the potential of an integrated data processing approach that combines untargeted and targeted elaboration (i.e., peak-region features, peaks, and visual features methods) to achieve highly effective fingerprinting from 2D patterns obtained by GC×GC and parallel dual secondary column-dual detection configuration. The work-flow, adapted from a previous study to combine untargeted and targeted investigation, is comprehensive yet efficient and fully supported by commercial software.

The sample set under study posed several analytical challenges related to the great compositional variability and complexity of the mice urine samples due to: (*a*) individual variations to diet manipulation; (*b*) timing of observation across the study; and (*c*) interfering chemicals from both derivatization and metabolic cages reservoirs. In this context, the combination of two parallel secondary columns was advantageous because this configuration results in higher ^{2}D column loadability and lower carrier gas velocities in the ^{2}D, thereby producing slightly wider and symmetrical peaks compatible with fast quadrupole detection. In addition, the dual detection provides complementary information: e.g. the MS provides essential information for identity confirmation and/or identification and supports highly specific untargeted cross-comparisons [13, 25, 30].

The *UT fingerprinting* data analysis work-flow is efficient because of the full integration of targeted (known compounds) with untargeted (peak-regions of detected 2D peaks) feature analysis, especially with exclusion of uninformative chromatogram regions (solvent peaks/strikes and column bleeding) or single 2D-peaks (derivatization artifacts) from the template and data processing. A detailed discussion on these aspects is provided in the following sections together with a brief interpretation of the experimental results. Insights into the evolution of individual metabolic signatures also are gleaned by combining *UT fingerprinting* and visual features investigation.

"UT fingerprinting" results

Mice urine samples analyzed in this study are listed in **Table 1** together with dietary compositions and collection times. The total number of runs was 108, with 3 diet groups × 6 mice-per-group × 3 collection points (timing).

The first step of the data elaboration work-flow aimed at chemical characterizing the urine profiles. Fifty three analytes, listed in **Table 2**, were reliably identified, including several aminoacids and related derivatives (glycine, hexanoylglycine, N-butyrylglycine, N-crotonylglycine, N-isovaleroylglycine, N-phenylacetylglycine), branched-chain amino acids - BCAA (leucine, isoleucine and valine), some carboxylic acids - TCA (cis-aconitic acid and citric acid) succinic acid, ,malic acid, α-ketoglutarate and fumaric acid, monosaccharides (fructose, glucose galactose, sucrose), and polyalcohols (myo-inositol, xylitol, glycerol, ribitol). Some analytes were of interest for the study because of their metabolic role is well defined and correlated with the dietary manipulation.

For example, hippuric acid is an acyl-glycine derivative; β-hydroxybutyric acid is involved in the synthesis and degradation of ketone bodies; and inorganic phosphate is related to the fructose phosphorylation status. BCAAs and their related metabolites have been demonstrated to correlate with development of insulin resistance (IR) in both rats and humans [31]. In the adipose tissue of patients with

obesity and type-2 diabetes mellitus (T2DM) with IR, the expression of genes encoding the enzymes of BCAA metabolism is significantly decreased through an undefined mechanism, leading to increased plasma levels of BCAAs, which persistently activate mTOR complex 1 (mTORC1), resulting in insulin resistance through the phosphorylation of insulin receptor substrate 1 (IRS-1) [32].

A first insight on the informative role of the selected known targets in describing the effects on the urinary metabolite fingerprint induced by the three dietary regimens was obtained by Principal Component Analysis on the targeted data and considering the Normalized 2D-Peak Responses (TIC-MS detection channel) as quantitative descriptors. Analytes that were not detected in some samples were included in the variables list and their response arbitrarily assigned at 5,000 counts corresponding to the 2D Peaks Volume detection threshold adopted for peak detection.

The PCA results on all targets × samples (53 × 108), illustrated in **Figure 2A** as scores plot, were connoted by a limited descriptive potential: the first two components (F1 - F2) accounted for only the 35% of the total variability without a clear independent sub-classification of diet regimens (control diet-CS, liquid fructose diet-FL, and solid fructose diet-FS). Still, most of samples derived from mice fed by liquid fructose solution were grouped in the IVth quadrant, as shown by confidence ellipses (green oval for p = 95%). CS group (blue tags) was less dispersed and co-clusterized with the FS group (purple tags). Possible explanations include the high biological variability, a limited descriptive potential of some targets included in the elaboration, and the presence of confounding variables.

Next, the set of targeted analytes was sieved to include just those with a statistically relevant variability between classes and/or those detected in just one sub-population. The criterion adopted was driven by the Fisher test; those with a $F_{calc} < F_{crit}$; with $F_{crit}(2, 6) = 5.14$ (α=0.5) were excluded from further computations.

Figure 2B illustrates PCA results based on the 37 most informative targets. The total variability explained by the first two components increases to 46% and the three groups are more clearly distinguishable. The CS and FS sub-populations reasonably overlap due to the expected lower systemic absorption of fructose (confirmed by serum metabolomics - data not shown). The lower bioavailability of fructose from FS diet seems to minimally impact the systemic metabolism derangements; however, the effect of higher fructose systemic distribution is evidenced by separately processed data from control (CS) and liquid fructose (FL) populations (**Figure 2C**). The group of FL individuals is differentiated from CS mice along F1 (accounting for the 32% of the total variability) while the timing of experiments is coherently distributed along F3 from bottom (FL T1) to the top (FLT 12). The first and the third principal components (F1 and F3) accounted here for about 42.4% of the total variance.

Within the group of most informative analytes distinguishing FL *vs.* CS, it is interesting to observe the contribution in the FL samples of the timing of the experiment (along F3 from bottom to top) of: (*a*) fructose

(with its two forms), glucose, galactose, ribitol, uric acid and threonine mostly contributing along F3 to discriminate week 1 samples against the others; (*b*) citric acid, pyruvic acid, malic acid, succinic acid and 2-keto glutaric acid, gluconic and cis-aconitic acid characterize 6 weeks of diet regimen samples; (*d*) malonic acid, glycine acyl derivatives (N-carboxy glycine, N-butyrylglycine, N-isovaleroylglycine, N-phenylacetylglycine) and hippuric acid closely related with 12 weeks of diet manipulation.

Univariate statistics for Normalized 2D-Peak Responses between sub-populations, visually illustrated by scatter diagrams of **Figure 3**, shows interesting trends. **Fig. 3A** shows fructose trend (the two detectable forms were both included and their normalized responses summed) between samples classes (e.g., CS, FL and FS).

Results for FL sub-population are connoted by a greater dispersion. An insight on FL sub-population suggests that there was a differential metabolic derangement along timing as a function of the dietary treatment. By mapping independently the normalized response of fructose isomers for FL samples at the three sampling time (e.g., 1, 6 and 12 weeks) a negative trend of excreted fructose is denoted.

Phenyl acetyl glycine is another analyte showing a characteristic trend between sub-classes. It contributes in characterizing samples collected at 12 weeks of liquid fructose feeding (FL T12) and its trend between sub-populations is connoted by a meaningful average difference between CS *vs.* FL and CS *vs.* FS (**Fig. 3B**). In fact, it is more abundant in the urine of mice fed by fructose independently by its supplemented form. In addition, its relative abundance in FL and FS sub-populations has relatively large dispersion (CV% was 50 and 63 respectively), most probably due to the kinetics of metabolic derangement. By observing the profile of phenyl acetyl glycine in the FL sub-population at 1, 6 and 12 weeks, the incremental trend over time is evident (**Fig. 3B** right end) and accompanied by a lower dispersion of results (CV% is 50% at T1, 22% at T6, and 18% at T12).

To see the effect of a solid fructose diet, a PCA was conducted by excluding the FL sub-population. The resulting scores plot is shown in **Figure 2D**. In this case, the differentiation between FS diet and controls is less evident although samples are distributed along F1 that explains the 25% of the total variance (i.e., 42.45%). In this case, the most informative variables were: glucose, hippuric acid (a glycine metabolite) and uric acid, more abundant in FS samples compared to CS; tartaric acid, exclusively present in the FS urines; and acyl glycine derivatives (N- carboxyl glycine, N-butyryl glycine and isovaleroil glycine), more abundant in the FS samples at the latest stage of diet manipulation (T 12 weeks). Fructose also was present in the group of most informative analytes (contributing 8% for the F1-explained variance).

The presence of higher glucose concentrations in FL and FS sub-populations, illustrated in **Figure 4**, could be explained by saturation of transporters as a consequence of serum distribution. Similarly, the high levels of glycine acyl derivatives is consistent with previous studies reporting glycine as biomarker of early diabetes onset [33]. Intriguingly, elevated glycine levels recently have been suggested to provide some

protection from fructose's deleterious effects. In fact, glycine has the potential to oppose the formation of Amadori products and Advanced Glycation Endproducts (AGEs. Further, through the activation of glycine-gated chloride channels present in intestinal L-cells and pancreatic α-cells, glycine may stimulate GLP-1 and glucagone release, which work in complementary ways to promote fatty acid oxidation and oppose fructose-evoked lipogenesis in the liver. With a similar mechanism, glycine antagonises Kupffer cell activation, which exacerbates fructose-induced steatosis. So, the increased levels of glycine and derivatives in individuals submitted to a fructose-enriched diet could be an attempt of the organism to defend itself from the fructose impact [34].

By extending the explorative investigation to all untargeted peak-regions covered by the *UT fingerprinting* (showed as red graphics in 2D chromatograms of **Figure 5A**), a cross-validation of previous results was obtained and further informative variables revealed. **Figure 6A** shows the results (score plot) of the PCA run by including untargeted peak-regions connoted by a meaningful variation across the set (F value > 5.14 - i.e., 266 over 380 peak-regions) from all individuals. Here, the combination of the first and the third principal components (F1 and F3) accounts for the 52% of the total variance; a better result if compared to the targeted analysis alone (**Figure 2B** - Total explained variance 42%). The more informative variables for the FL sub-set against the control (CS) and the solid fructose diet (FS) are: feature #98 (25.37min - 1.89 s), corresponding to pyroglutamic acid; feature #540 (39.66 min - 1.35s), malic acid; feature #50 (26.53 min - 1.02 s), threonic acid; feature #48 (37.23 min - 1.02 s), mannitol; feature #649 (38.15 min - 2.20 s), gluconic acid; feature #245 (36.06 min - 1.82 s), fructose; and feature #33 (36.26 min - 1.00 s), corresponding to one of the glucose derivatives. All these analytes, ranked in the first thirty most informative peak-regions (% of variable contribution along F1), were already mapped by the targeted fingerprinting thereby their role resulted cross-validated.

Some additional features revealed by the untargeted approach were not previously (*a priori*) targeted analysis, but indicate possible additional information that could more completely characterize the phenomenon under study. Features #97 and #200 (41.03 min - 1.56 s), corresponding to the N-acetyl glucosamine retention region; features #292 (29.34 min - 2.14 s) and #528 (33.17 min - 2.93s), possibly glycine derivatives tentatively identified as malonyl glycine and methyl malonyl glycine (1570 and 1695 I^T_S respectively); feature #331 (36.5 min - 3.14 s), possibly N-acetyl glutamine (1748 I^T_S); and feature #219 (21.76 min - 1.36 s), glutaric acid (1269 I^T_S).

A further confirmation not only on (*a*) the relevance of the monitored targeted analytes in describing the metabolic derangement, but also on (*b*) the necessity of extending the list of known analytes to better characterize the phenomenon is evident by observing the PCA results when peak-region features from CS and FL sub-populations were analyzed (**Figure 6B**). The FL population is again more dispersed on the Cartesian plane compared to the control samples, but it is coherently distributed along F3 with experiment timing as in

the case of targeted elaboration (also in the PCA of **Fig. 2C**). In addition, the first Principal Component (F1) describes a higher portion of the total variance (51.9 % F1 **Fig. 6B** *vs*. 32.2% of **Fig. 2C**).

As far as the additional analytes are concerned, N-acetyl glucosamine was 70 (FL) and 30 (FS) times higher in dietary-manipulated groups than in the control group. A similar trend was observed for methyl-malonylglycine, which was 12 (FL) and 8 (FS) times higher than in the CS group; malonylglycine showed an average a fourfold increase, of statistical relevance, only in the FL group, while glutaric acid was 2.5 (FL) and 2 (CS) times higher than in FS group.

Targeted quantitative fingerprinting

Quantitative analysis was run by 1D GC-MS on *a priori* selection of target analytes implemented with some informative metabolites indicated by the *UT fingerprinting*. The list includes compounds that were correlated with a more drastic dietary insult [25], early biomarkers of prediabetic/insulin resistance phenotypes as well as related hepatic and renal dysfunctions.

Quantitative data affords reliable comparisons and help in evaluating possible deviations from physiological or reference values independently from the analytes response factors and system dynamic range. Highly abundant or variably distributed analytes, such as some of the monosaccharides (glucose, fructose and galactose), when accurately quantified, reveal informative trends along experiment timing and between sub-populations. Analyte distributions in the samples are visualized as heat-maps (based on analytes concentration in mg/L of urine - data are displayed as absolute concentrations (SF2a) or normalized concentration *vs.* creatinine content (SF2b)) and provided as Supplementary material (Supplementary Figures 2a and 2b). Hierarchical clustering based on Euclidean distances was also performed to highlight those analytes with an higher informative role in describing the metabolic derangement. As illustrated in SF2a and SF2b raw and normalized data have comparable information potential as their clustering is the same. Some insights on single metabolites distribution will follow.

Fructose, for example, was 11 times more abundant in the FL group than in the control (CS) while in the FS diet group it showed a 4 fold increase. GC×GC data, on the other hand, exhibited increases of 9 fold (FL vs. CS) and 4 fold (FS vs. CS), in agreement with quantitative data. Hippuric acid, was 3.2 (FL) and 1.3 (FS) times higher than in the control, confirming the trend observed by GC×GC, while citric acid showed a 1.5 fold increase, of statistical relevance just in the FL diet group.

To cross-validate fingerprinting results (response data by targeted GC×2GC-MS/FID and quantitative data by GC-MS), a Multiple Factor Analysis (MFA) on the targeted peaks distribution was performed. MFA makes it possible to analyze two data matrix/tables simultaneously and provides results capable of assessing

the relationship between the observations (sample sub-classes) and the variables (targeted analytes). For each data matrix/table, PCA is performed first, then the value of the first eigenvalue of each analysis is stored and used to weight the two matrix/tables in a second step of the analysis. Results, expressed as RV coefficients, indicate to what extent the tables/variables distributions are related two-by-two. The more variables from the GC×2GC-MS/FID method are related to the GC-MS quantitative counterpart, the higher the RV coefficient (variation range 0-1). In this case, the mutual correlation between the two data sets was 0.854, indicating a good overall consistency between fingerprints informative content.

More intuitively, the MFA map of the projected points, shown in **Figure 7** and including control (CS) and liquid fructose (FL) individuals illustrates sample distributions resulting from the information provided by the two data-sets. Lines connect each single observation to its relative position (projection) on the two independent PCA charts (GC-MS method - "Table 1" and GC×2GC-MS/FID method - "Table 2"). The projected points correspond to supplementary observations for which only the information provided by one table is taken into account, the other tables being transformed to 0s. The chart based on CS and FL diet groups, where combining the first two PC (F1 and F2) the total explained variance raises the 43.38 %, the two sub populations are clearly distinguishable. .

Urine fingerprint evolution by visual features analysis

The last part of this study focuses on a fingerprinting approach based on visual features. Visualization suitable for rapid and effective pair-wise pattern comparisons, giving the possibility of monitoring variations during experimentation from single individuals as in the personalized medicine applications [9]. Of great interest in this direction is the possibility to monitor metabolic changes in a single subject to better staging the disease.

Visual features fingerprinting, also defined as image comparison, is one of the earliest introduced in GC×GC data elaboration[10, 35] and is very effective to capture distinctive 2D patterns, compare them on an untargeted basis to promptly reveal compositional differences [18, 25, 36–38] The approach is fully automated, and when applied after *UT fingerprinting*, provides information about targeted or untargeted peak-regions variations between pair-wise compared samples.

The example of **Figure 5** refers to Mouse #40 urine samples obtained at 12 weeks (**Figure 5A**) of dietary regimen with liquid fructose supplementation (averaged normalized image of two analytical replicates) considered as the *analyzed* image versus the urine samples collected at 6 weeks (**Figure 5B**) (averaged normalized image of two analytical replicates) considered as the *reference* image. The resulting image (**Figure 5C**) is rendered as "colorized fuzzy ratio"; the difference at each data point between aligned pair-wise images is

computed and colored green, when positive (larger detector response in the analyzed image (T12)) or colored red, when the difference is negative (larger detector response in the reference image (T6)). Brightness depends on the magnitude of the responses, white saturation indicates pixels/regions with detector responses nearly equal in pair images.

Previously processed 2D chromatograms submitted to the image comparison already were mapped by the peak-regions template, with the 2D peaks identity (if known) or unique identification numbering (#) for unknowns included as metadata for visual tracking of differences. Informative variations (positive trends and negative trends) are registered for analytes listed in **Figure 5C** (Normalized 2D Peak Volume variations are also reported) and visually tracked on the colorized differential image with yellow/red contour lines.

Fingerprint variations are perfectly coherent with the trends observed for the entire sub-population: fructose (here mapped in its two derivatized forms), 2-keto glutaric, citric, succinic, and malic and lactic acid all showing a negative trend along experiments timing from 6 to 12 weeks, confirming the evidence of a homestatic adaptation to the dietary insult. On the other hand, analytes markedly increased in the week 12 samples are acyl glycine derivatives (N-crotonyl glycine, N-isovaleroyl glycine, N-phenyl acetyl glycine N-carboxy-glycine and hexanoyl glycine), L-threonic acid, L-threonine, and glucose.

Visual features methods provide a direct and prompt approach to localize analytes with large variations between pairs, confirming (or not) the individual trends within the sub-populations, but, at the same time, highlighting specific metabolism alterations.

Conclusions

Experimental results obtained by an advanced fingerprinting approach to study the evolution of the urinary metabolic signature in mice under dietary manipulation confirmed some of the expectations and provided useful information to understand better the systemic effect of the dietary differences.

In particular:

1) *UT fingerprinting* on 2D data obtained by GC×2GC-MS/FID cross-validates the informative role of a series of small metabolites excreted with urine in mice subjected to an high fructose intake (solid or liquid form) compared to those fed by a control diet. Metabolites whose quantitative variations were considered statistically relevant are: fructose, glucose, glycine and some of its acyl-derivatives (N-carboxyglycine, N-butyrylglycine, N-isovaleroylglycine, N-phenylacetylglycine), uric acid, hippuric acid, citric acid, threonic acid, malic acid, succinic acid, 2-ketoglutaric acid, and tartaric acid.

2) The *UT fingerprinting* based on peak-regions extended the comparative process to a higher number of features while confirming the targeted analysis result. In addition, additional insights on the phenomenon were obtained with new targets highlighted as discriminant: two glycine acyl-derivatives (malonylglycine and methyl-

malonylglycine), N-acetylglutamine, and glutaric acid. These are highly informative compounds, with roles connected with fructose metabolism, so they merit future in-depth studies.

3) Individual variations (within groups and along the timing of the experiments) suggested the adoption of visual features approaches to promptly monitor metabolites variations. The results of image comparison on a single mouse for urines collected after six and twelve weeks were in line with those derived from the global population, but informed on the effect of the diet on a single individual.

4) Quantitative analysis obtained by external calibration and GC-MS confirmed the trends observed by GC×GC-MS/FID, providing consistent and reliable information on the real impact of dietary manipulation. Quantitative data on metabolites excretion could be the starting point of a complete and interdisciplinary evaluation of the real impact of the diet in the mice physiological status.

Perhaps most notably, our results lead to the identification of different urine metabolite fingerprints recorded for the two fructose groups. These findings are very important as they confirm and further extend on-going observations on the different toxicological impacts of solid vs liquid fructose supplementation. Specifically, the FL urine samples consistently were differentiated and clusterized independently from the others, suggesting that metabolic alterations are more significant. Overall, these data suggest that the approach here proposed could be used for better stratifications of populations which at first glance seem to be exposed to the same metabolic insult (fructose overconsumption) but their metabolic profiles are affected by slight differences in the dietary manipulation (here solid vs. liquid form).

Further studies are required to test whether the selected metabolites might be useful as diagnostic tools and to reveal the biological mechanisms that result in changes in the levels of certain metabolites in the pathogenesis of fructose- induced metabolic impairments.

Acknowledgements

Authors are indebted with Dr. Martina Carbone for the technical help.

This work was supported and funded by the University of Turin (Ricerca Locale 2016 and 2017), Regione Toscana (Bando Nutraceutica 2014), Project TAGIDISFRU, ERA-HDHL-ERANET Biomarkers for Nutrition and Health Implementing the JPG HDHL objectives" - Project "SALIVAGES".

Compliance with Ethical Standards

Authors declare no conflict of interests.

The research was conducted in accordance with the European Directive 2010/63/EU on the protection of animals used for scientific purposes as well as the International Guiding Principles for Biomedical Research

Involving Animals, issued by the Council for the International Organizations of Medical Sciences. Both the Turin University Ethics Committee (Prot. 6437 22/02/2016) and the Italian Ministry of Health (Prot. 42/2017-PR) approved the experimental protocol. All efforts were made to minimize animal suffering and to reduce the number of animals used.

References

1. Malik VS, Hu FB (2015) Fructose and Cardiometabolic Health What the Evidence from Sugar-Sweetened Beverages Tells Us. J Am Coll Cardiol 66:1615–1624. doi: 10.1016/j.jacc.2015.08.025
2. Rizkalla SW (2010) Health implications of fructose consumption: A review of recent data.(Review)(Report). Nutr Metab (Lond) 7:82. doi: 10.1186/1743-7075-7-82
3. Lim JS, Mietus-Snyder M, Valente A, Schwarz J-M, Lustig RH (2010) The role of fructose in the pathogenesis of NAFLD and the metabolic syndrome. Nat Rev Gastroenterol Hepatol 7:251–264.
4. Teff KL, Grudziak J, Townsend RR, Dunn TN, Grant RW, Adams SH, Keim NL, Cummings BP, Stanhope KL, Havel PJ (2009) Endocrine and Metabolic Effects of Consuming Fructose- and Glucose-Sweetened Beverages with Meals in Obese Men and Women: Influence of Insulin Resistance on Plasma Triglyceride Responses. J Clin Endocrinol Metab 94:1562–1569. doi: 10.1210/jc.2008-2192
5. Lê KA, Ith M, Kreis R, Faeh D, Bortolotti M, Tran C, Boesch C, Tappy L (2009) Fructose overconsumption causes dyslipidemia and ectopic lipid deposition in healthy subjects with and without a family history of type 2 diabetes. Am J Clin Nutr 89:1760–1765. doi: 10.3945/ajcn.2008.27336
6. Schwarz JM, Noworolski SM, Wen MJ, Dyachenko A, Prior JL, Weinberg ME, Herraiz LA, Tai VW, Bergeron N, Bersot TP, Rao MN, Schambelan M, Mulligan K (2015) Effect of a high-fructose weight-maintaining diet on lipogenesis and liver fat. J Clin Endocrinol Metab 100:2434–2442. doi: 10.1210/jc.2014-3678
7. Ritze Y, Bárdos G, D'Haese JG, Ernst B, Thurnheer M, Schultes B, Bischoff SC (2014) Effect of high sugar intake on glucose transporter and weight regulating hormones in mice and humans. PLoS One 9:1–9. doi: 10.1371/journal.pone.0101702
8. Mastrocola R, Ferrocino I, Liberto E, Chiazza F, Cento AS, Querio G., Nigro D, Bitonto V, Cutrin JC, Rantsiou K, Durante M, Masini E, Aragno M, Cordero C, Cocolin L CM Fructose liquid and solid formulations differently affect gut integrity, microbiota composition and related liver toxicity: a comparative in vivo study. J. Biochem. Nutr.
9. Hunter P (2009) Reading the metabolic fine print. EMBO Rep 10:20–23. doi: 10.1038/embor.2008.236
10. Almstetter MF, Oefner PJ, Dettmer K (2012) Comprehensive two-dimensional gas chromatography in metabolomics. Anal Bioanal Chem 402:1993–2013. doi: 10.1007/s00216-011-5630-y
11. Reichenbach SE, Tian X, Tao Q, Ledford EB, Wu Z, Fiehn O (2011) Informatics for cross-sample analysis with comprehensive two-dimensional gas chromatography and high-resolution mass spectrometry (GCxGC-HRMS). Talanta 83:1279–1288. doi: 10.1016/j.talanta.2010.09.057
12. Koek MM, Muilwijk B, van Stee LLP, Hankemeier T (2008) Higher mass loadability in comprehensive two-dimensional gas chromatography-mass spectrometry for improved analytical performance in

metabolomics analysis. J Chromatogr A 1186:420–429. doi: 10.1016/j.chroma.2007.11.107

13. Nicolotti L, Cordero C, Bressanello D, Cagliero C, Liberto E, Magagna F, Rubiolo P, Sgorbini B, Bicchi C (2014) Parallel dual secondary column-dual detection: A further way of enhancing the informative potential of two-dimensional comprehensive gas chromatography. J Chromatogr A 1360:264–274. doi: 10.1016/j.chroma.2014.07.081
14. Reichenbach SE, Rempe DW, Tao Q, Bressanello D, Liberto E, Bicchi C, Balducci S, Cordero C (2015) Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors. Anal Chem 87:10056–10063. doi: 10.1021/acs.analchem.5b02718
15. Sgorbini B, Cagliero C, Boggia L, Liberto E, Reichenbach SE, Rubiolo P, Cordero C, Bicchi C (2015) Parallel dual secondary-column-dual detection comprehensive two-dimensional gas chromatography: a flexible and reliable analytical tool for essential oils quantitative profiling. Flavour Fragr J 30:366–380. doi: 10.1002/ffj.3255
16. Cordero C, Rubiolo P, Cobelli L, Stani G, Miliazza A, Giardina M, Firor R, Bicchi C (2015) Potential of the reversed-inject differential flow modulator for comprehensive two-dimensional gas chromatography in the quantitative profiling and fingerprinting of essential oils of different complexity. J Chromatogr A 1417:79–95. doi: 10.1016/j.chroma.2015.09.027
17. Cordero C, Rubiolo P, Reichenbach SE, Carretta A, Cobelli L, Giardina M, Bicchi C (2016) Method translation and full metadata transfer from thermal to differential flow modulated comprehensive two dimensional gas chromatography: Profiling of suspected fragrance allergens. J Chromatogr A. doi: 10.1016/j.chroma.2016.12.011
18. Magagna F, Valverde-Som L, Ruíz-Samblás C, Cuadros-Rodríguez L, Reichenbach SE, Bicchi C, Cordero C (2016) Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil. Anal Chim Acta 936:245–258. doi: 10.1016/j.aca.2016.07.005
19. Zhang Q, Wang G, Du Y, Zhu L, A J (2007) GC/MS analysis of the rat urine for metabonomic research. J Chromatogr B Anal Technol Biomed Life Sci 854:20–25. doi: 10.1016/j.jchromb.2007.03.048
20. Sgorbini B, Cagliero C, Boggia L, Liberto E, Reichenbach SE, Rubiolo P, Cordero C, Bicchi C (2015) Parallel dual secondary-column-dual detection comprehensive two-dimensional gas chromatography: a flexible and reliable analytical tool for essential oils quantitative profiling. Flavour Fragr J 30:366–380. doi: 10.1002/ffj.3255
21. Reichenbach SE, Carr PW, Stoll DR, Tao Q (2009) Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography. 1216:3458–3466. doi: 10.1016/j.chroma.2008.09.058

22. Sgorbini B, Cagliero C, Boggia L, Liberto E, Reichenbach SE, Rubiolo P, Cordero C, Bicchi C (2015) Parallel dual secondary-column-dual detection comprehensive two-dimensional gas chromatography: A flexible and reliable analytical tool for essential oils quantitative profiling. Flavour Fragr J. doi: 10.1002/ffj.3255
23. Reichenbach SE, Tian X, Boateng AA, Mullen CA, Cordero C, Tao Q (2013) Reliable peak selection for multisample analysis with comprehensive two-dimensional chromatography. Anal Chem 85:4974–4981. doi: 10.1021/ac303773v
24. Reichenbach SE, Tian X, Cordero C, Tao Q (2012) Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography. J Chromatogr A 1226:140–148. doi: 10.1016/j.chroma.2011.07.046
25. Bressanello D, Liberto E, Collino M, Reichenbach SE, Benetti E, Chiazza F, Bicchi C, Cordero C (2014) Urinary metabolic fingerprinting of mice with diet-induced metabolic derangements by parallel dual secondary column-dual detection two-dimensional comprehensive gas chromatography. J Chromatogr A 1361:265–276. doi: 10.1016/j.chroma.2014.08.015
26. Hollingsworth B V., Reichenbach SE, Tao Q, Visvanathan A (2006) Comparative visualization for comprehensive two-dimensional gas chromatography. pp 51–58
27. Eurachem (2014) Eurachem guide: the fitness for purpose of analytical methods – a laboratory guide to method validation and related topics.
28. Citac, Eurachem (2000) Quantifying Uncertainty in Analytical Measurement. English 2nd:126. doi: 0 948926 15 5
29. Commission E (2002) Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. Off J Eur Union L221:8–36.
30. Reichenbach SE, Rempe DW, Tao Q, Bressanello D, Liberto E, Bicchi C, Balducci S, Cordero C (2015) Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors. Anal Chem 87:10056–10063. doi: 10.1021/acs.analchem.5b02718
31. Zhao X, Han Q, Liu Y, Sun C, Gang X, Wang G (2016) The Relationship between Branched-Chain Amino Acid Related Metabolomic Signature and Insulin Resistance: A Systematic Review. J Diabetes Res. doi: 10.1155/2016/2794591
32. Fontana L, Cummings NE, Apelo SIA, Joshua C, Kasza I, Schmidt BA, Cava E, Spelta F, Syed FA, Baar EL, Veronese N, Cottrell SE, Fenske RJ, Bertozzi B, Brar HK, Pietka T, Arnold D, Figenshau RS, Andriole GL, Merrins MJ, Alexander M, Kimple ME, Lamming DW, Sciences N, Sciences E, Avanzate B, Memorial SM, Hospital V (2016) HHS Public Access. Cell Rep 16:520–530. doi: 10.1016/j.celrep.2016.05.092.Decreased
33. Tulipani S, Palau-Rodriguez M, Mi??arro Alonso A, Cardona F, Marco-Ramell A, Zonja B, Lopez de Alda

M, Mu??oz-Garach A, Sanchez-Pla A, Tinahones FJ, Andres-Lacueva C (2016) Biomarkers of Morbid Obesity and Prediabetes by Metabolomic Profiling of Human Discordant Phenotypes. Clin Chim Acta 463:53–61. doi: 10.1016/j.cca.2016.10.005

34. McCarty MF, DiNicolantonio JJ (2014) The cardiometabolic benefits of glycine: Is glycine an “antidote” to dietary fructose? Open Hear 1:e000103. doi: 10.1136/openhrt-2014-000103
35. Reichenbach SE, Tian X, Cordero C, Tao Q (2012) Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography. J Chromatogr A 1226:140–148. doi: 10.1016/j.chroma.2011.07.046
36. Cordero C, Zebelo SA, Gnavi G, Griglione A, Bicchi C, Maffei ME, Rubiolo P (2012) HS-SPME-GC×GC-qMS volatile metabolite profiling of Chrysolina herbacea frass and Mentha spp. leaves. Anal Bioanal Chem 402:1941–1952. doi: 10.1007/s00216-011-5600-4
37. Cordero C, Liberto E, Bicchi C, Rubiolo P, Reichenbach SE, Tian X, Tao Q (2010) Targeted and non-targeted approaches for complex natural sample profiling by GCxGC-qMS. J Chromatogr Sci 48:251–261. doi: 10.1093/chromsci/48.4.251
38. Purcaro G, Cordero C, Liberto E, Bicchi C, Conte LS (2014) Toward a definition of blueprint of virgin olive oil by comprehensive two-dimensional gas chromatography. J Chromatogr A 1334:101–111. doi: 10.1016/j.chroma.2014.01.067

Tables Caption

Table 1: diet details, number of animals per group, timing of experiments and number of analyses.

Table 2: list of targeted compounds from GC×2GC-MS/FID analysis; compounds are listed together with retention times (^1t_R ; ^2t_R), intermediate precision (relative standard deviation - RSD%) estimated on 108 runs, experimental and tabulated linear retention indexes (I^T_S).

Figure Captions

Figure 1: 2D data elaboration workflow combining targeted and untargeted investigation

Figure 2: Principal component analysis (PCA) scores plot referred to: (**2A**) all targeted analytes distribution across 108 runs; (**2B**) targeted analytes with a statistically relevant variability between classes and Fisher values above 5.14; (**2C**) sub-set of individuals fed by control diet (CS) and liquid fructose (FL); (**2D**) sub-set of individuals fed by control diet (CS) and solid fructose (FL).

Figure 3: Univariate statistics (by scatter diagrams) on 2D normalized volumes between sub-populations: (**3A**) fructose trend between populations (left side) and (right side) within the liquid fructose (FL) group along timing (1,6,12 weeks); (**3B**) Phenyl acetyl glycine trends populations (left side) and (right side) within the liquid fructose (FL) group along timing (1,6,12 weeks).

Figure 4: glucose normalized responses between diet populations and experiment timings.

Figure 5: 2D patterns (MS signal- Full scan acquisition) of urine samples collected at 12 weeks - *analyzed* image (5A) and at 6 weeks - *reference* image (5B) from Mouse #40 fed with liquid fructose supplementation. Red areas located untargeted peak-regions delineated by the UT fingerprinting work-flow. Green areas/peak-regions were excluded by the untargeted elaborations since they correspond to column bleeding and/or derivatization reagents interferences. Comparative visualization (5C) is rendered as "colorized fuzzy ratio" and reveals compositional differences (normalized responses) between the two urine samples. Targeted compounds with meaningful variations are listed on the right end of the image together with the relative response differences.

Figure 6: Principal component analysis (PCA) scores plot referred to: (**6A**) normalized 2D volumes of all untargeted peak-regions connoted by a meaningful variation across the set (F value > 5.14 - i.e., 266 over 380 peak-regions) and from all individuals; (**6B**) refers to control diet and liquid fructose populations only.

Figure 7: Chart of the projected points obtained by Multiple Factor Analysis (MFA) on the targeted peaks distribution as obtained by GC-MS accurate quantitation (Table 1 in samples' tag) and by GC×2GC-MS/FID as semi-quantitative profiling method (Table 2 in samples' tag). Lines connect each single observation to its relative position (projection) on the two independent PCA charts. The chart refers to CS and FL diet groups.

Figure 1

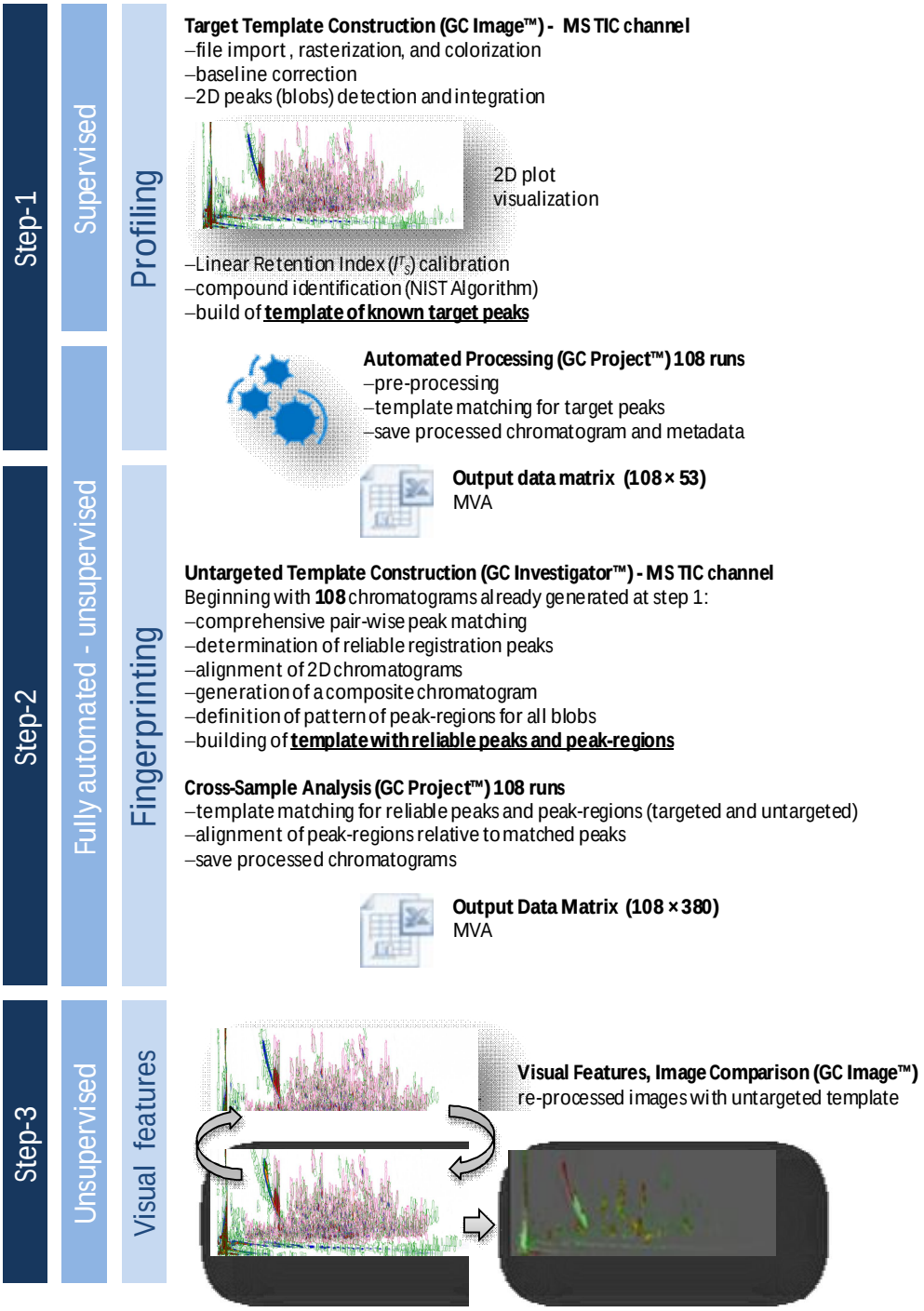

Step-1
Supervised
Profiling
Target Template Construction (GC Image™) - MS TIC channel
–file import , rasterization, and colorization
–baseline correction
–2D peaks (blobs) detection and integration
2D plot visualization
–Linear Retention Index (I^T_S) calibration
–compound identification (NIST Algorithm)
–build of template of known target peaks
Automated Processing (GC Project™) 108 runs
–pre-processing
–template matching for target peaks
–save processed chromatogram and metadata
Output data matrix (108 × 53)
MVA
Step-2
Fully automated - unsupervised
Fingerprinting
Untargeted Template Construction (GC Investigator™) - MS TIC channel
Beginning with 108 chromatograms already generated at step 1:
–comprehensive pair-wise peak matching
–determination of reliable registration peaks
–alignment of 2D chromatograms
–generation of a composite chromatogram
–definition of pattern of peak-regions for all blobs
–building of template with reliable peaks and peak-regions
Cross-Sample Analysis (GC Project™) 108 runs
–template matching for reliable peaks and peak-regions (targeted and untargeted)
–alignment of peak-regions relative to matched peaks
–save processed chromatograms
Output Data Matrix (108 × 380)
MVA
Step-3
Unsupervised
Visual features
Visual Features, Image Comparison (GC Image™)
re-processed images with untargeted template

Figure 2

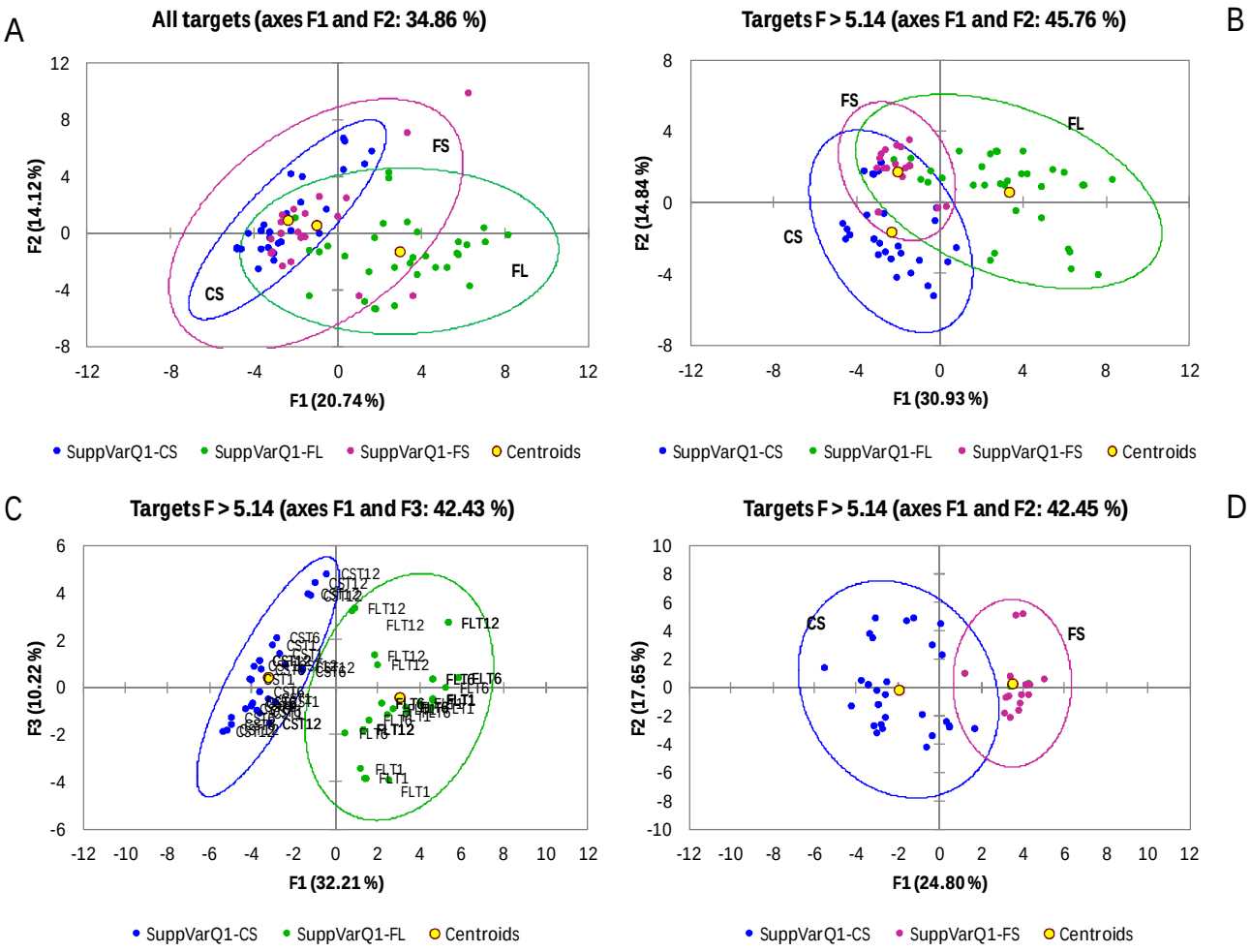

A
All targets (axes F1 and F2: 34.86 %)
F2 (14.12 %)
F1 (20.74 %)
FS
FL
CS
SuppVarQ1-CS
SuppVarQ1-FL
SuppVarQ1-FS
Centroids
B
Targets F > 5.14 (axes F1 and F2: 45.76 %)
F2 (14.84 %)
F1 (30.93 %)
FS
FL
CS
SuppVarQ1-CS
SuppVarQ1-FL
SuppVarQ1-FS
Centroids
C
Targets F > 5.14 (axes F1 and F3: 42.43 %)
F3 (10.22 %)
F1 (32.21 %)
SuppVarQ1-CS
SuppVarQ1-FL
Centroids
D
Targets F > 5.14 (axes F1 and F2: 42.45 %)
F2 (17.65 %)
F1 (24.80 %)
CS
FS
SuppVarQ1-CS
SuppVarQ1-FS
Centroids

Figure 3

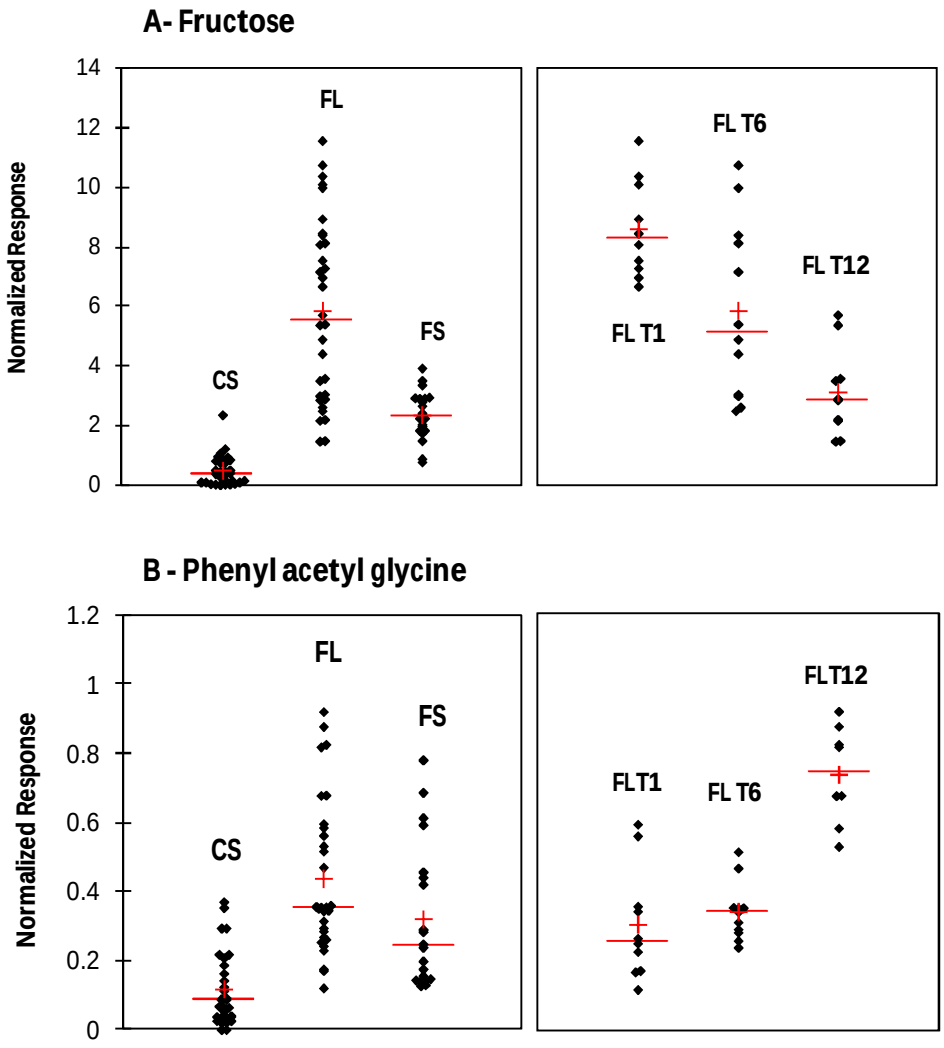
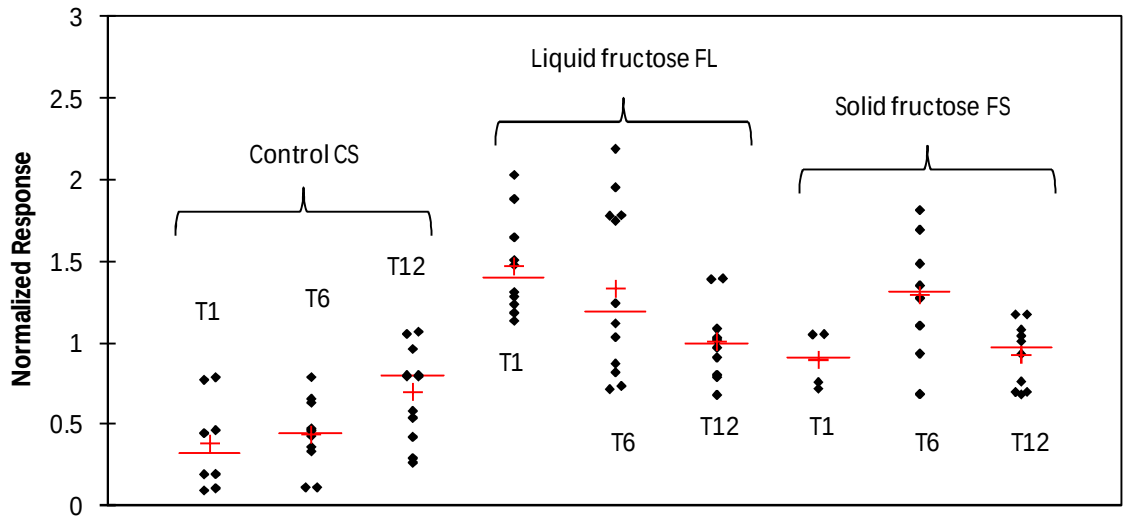

A- Fructose
Normalized Response
14
12
10
8
6
4
2
0
CS
FL
FS
FL T1
FL T6
FL T12
B - Phenyl acetyl glycine
Normalized Response
1.2
1
0.8
0.6
0.4
0.2
0
CS
FL
FS
FLT1
FL T6
FLT12

Figure 4


Normalized Response
3
2.5
2
1.5
1
0.5
0
Control CS
Liquid fructose FL
Solid fructose FS
T1
T6
T12
T1
T6
T12
T1
T6
T12

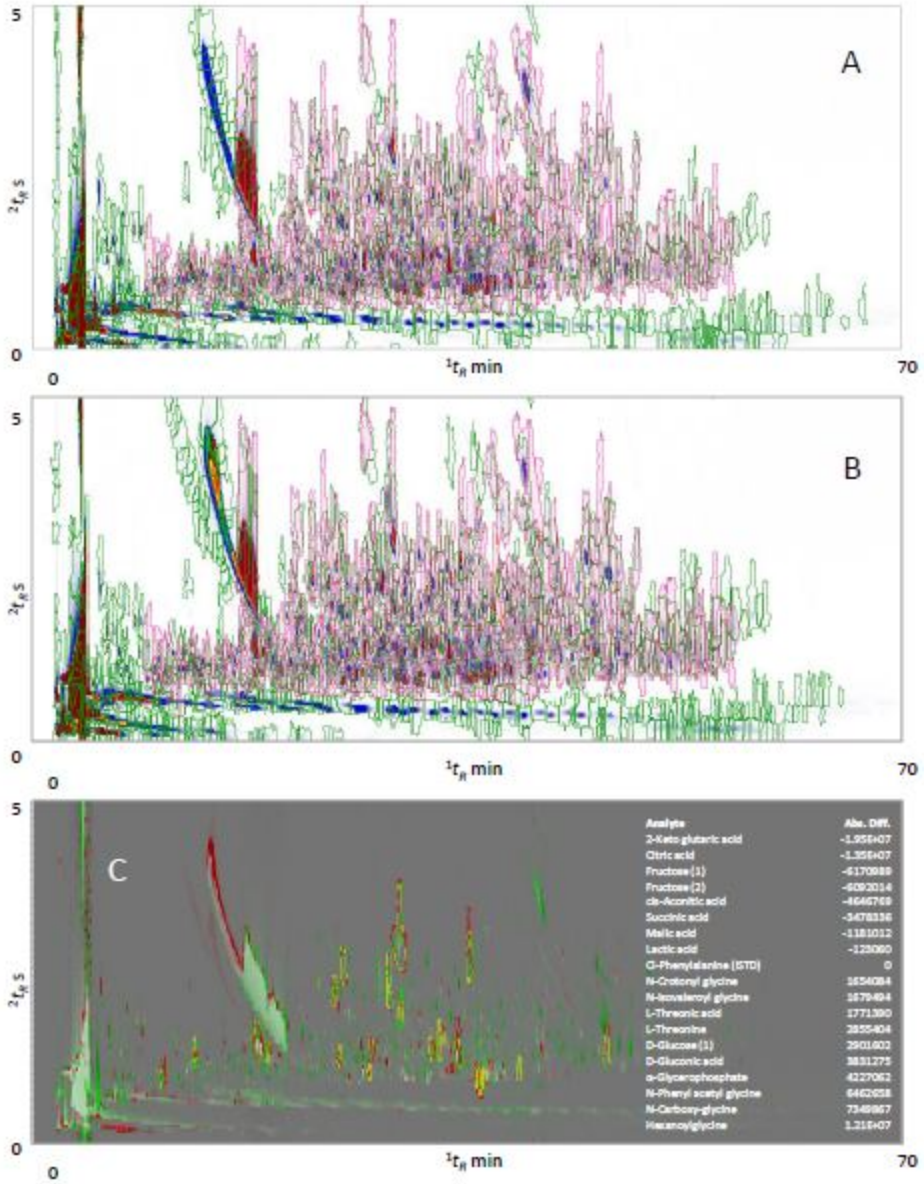

A
B
C
5
0
70
^2t_R s
^1t_R min
Analyte
Abs. Diff.
2-Keto glutaric acid -1.95E+07
Citric acid -1.35E+07
Fructose (1) -6170989
Fructose (2) -6092014
cis-Aconitic acid -4646769
Succinic acid -3478336
Malic acid -1181012
Lactic acid -123060
Cl-Phenylalanine (STD) 0
N-Crotonyl glycine 1654084
N-Isovaleroyl glycine 1679494
L-Threonic acid 1771390
L-Threonine 2855404
D-Glucose (1) 2901602
D-Gluconic acid 3831275
α-Glycerophosphate 4227062
N-Phenyl acetyl glycine 6462658
N-Carboxy-glycine 7348967
Hexanoylglycine 1.21E+07

Figure 6

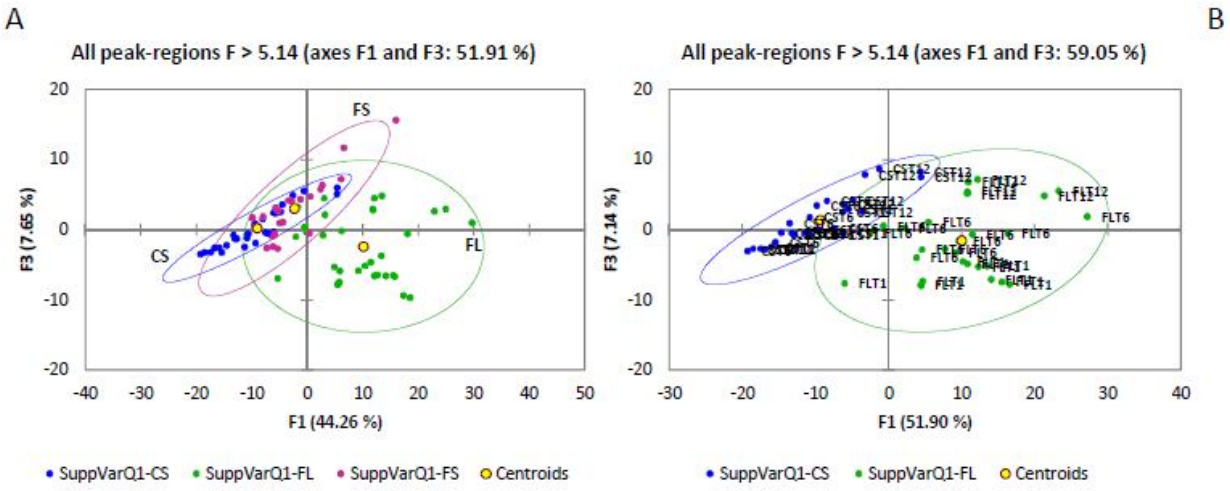

A
All peak-regions F > 5.14 (axes F1 and F3: 51.91 %)
F3 (7.65 %)
F1 (44.26 %)
FS
CS
FL
SuppVarQ1-CS
SuppVarQ1-FL
SuppVarQ1-FS
Centroids
B
All peak-regions F > 5.14 (axes F1 and F3: 59.05 %)
F3 (7.14 %)
F1 (51.90 %)
SuppVarQ1-CS
SuppVarQ1-FL
Centroids

Figure 7

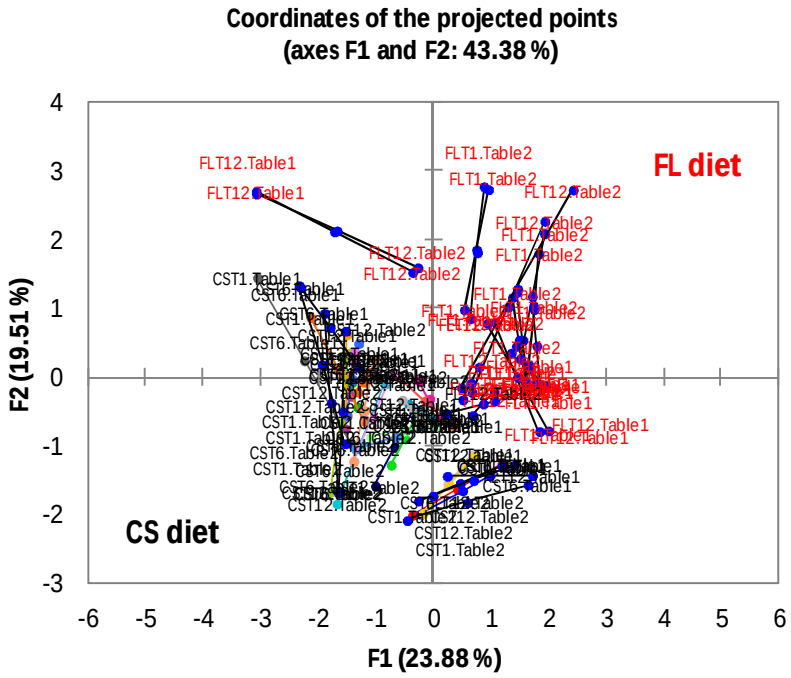

Coordinates of the projected points
(axes F1 and F2: 43.38 %)
FL diet
CS diet
F2 (19.51 %)
F1 (23.88 %)
FLT12.Table1
FLT1.Table2
FLT12.Table2
CST1.Table1
CST6.Table1
CST12.Table2
CST1.Table2

Table 1

Diet Group	Diet Energy balance calories %	n° of animals	Timing weeks	Analytical replicates	Number of analytial runs
Control (CS)	– carbohydrates 62.1% – [36.2% starch, 14.9% dextrin and 11% sugar] – fats 5.1% – fibres 5% – proteins 17.6% Drinking tap water	6	1(basal) 6,12	2	36
Liquid fructose (FL)	– carbohydrates 70% – [8.9% starch, 1% sugar, 60% fructose liquid solution] – fats 10% – proteins 20% Drinking fructose solution (60% w/w and citric acid as additive)	6	1(basal) 6,12	2	36
Solid fructose (FS)	– carbohydrates 70% – [8.9% starch, 1% sugar, 60% fructose solid form] – fats 10% – proteins 20% Drinking tap water	6	1(basal) 6,12	2	36

Table 2

Analyte	^{1}D Retention time ^1t_R (min) Mean	Stdev	RSD	^{2}D Retention time ^2t_R (sec) Mean	Stdev	RSD	experimental I^T_S	reference I^T_S
$Pyruvic acid	9.99	0.12	1.15%	1.04	0.01	1.02%	1059	1099
$Lactic acid	10.36	0.11	1.04%	0.97	0.02	1.59%	1067	1066
$L-Alanine	11.72	0.13	1.09%	0.98	0.03	2.79%	1110	1095
N-carboxy-glycine	13.01	0.10	0.79%	1.27	0.03	2.40%	1147	-
$L-Isoleucine	13.48	0.28	2.06%	1.26	0.04	3.50%	1158	1286
$3-Hydroxybutyric acid	13.61	0.11	0.79%	1.01	0.02	2.41%	1168	1148
$2-Keto-3-methyl-N-valeric acid	14.36	0.10	0.71%	1.14	0.02	1.60%	1189	1206
$Malonic acid	15.19	0.10	0.68%	1.32	0.02	1.44%	1218	1216
$L-Valine	15.63	0.06	0.38%	1.00	0.00	0.00%	1227	1224
$L-Leucine	17.56	0.10	0.57%	1.00	0.01	1.21%	1273	1272
$Glycerol	17.79	0.09	0.48%	0.87	0.03	3.27%	1292	1289
Phosphate	18.09	0.15	0.83%	1.41	0.04	2.59%	1299	1290
$Glycine	18.65	0.11	0.57%	1.03	0.01	1.16%	1318	1110
$Succinic acid	18.86	0.09	0.47%	1.34	0.03	2.27%	1322	1321
$Fumaric acid	19.85	0.09	0.47%	1.30	0.02	1.56%	1355	1353
2,3-Dihydroxybutanoic acid	20.42	0.10	0.47%	1.04	0.03	2.61%	1366	1439
$L-Threonine	21.43	0.07	0.30%	1.04	0.00	0.00%	1395	1377
$L-Aspartic acid	22.34	0.00	0.00%	1.54	0.00	0.00%	1418	1422
N-Butyryl glycine	22.92	0.08	0.36%	2.37	0.05	2.07%	1431	1505
N-Isovaleroylglycine	24.23	0.09	0.37%	2.26	0.03	1.45%	1467	1502
$Malic acid	24.63	0.07	0.29%	1.22	0.03	2.06%	1480	1478
N-Crotonylglycine	24.88	0.08	0.33%	2.51	0.04	1.74%	1493	1482**
Pyroglutamic acid	25.38	0.08	0.31%	1.90	0.04	1.96%	1526	1523
L-Threonic acid	27.08	0.08	0.28%	1.03	0.03	3.39%	1579	1545
$2-keto glutaric acid	27.20	0.08	0.28%	1.45	0.06	3.89%	1581	1572
Hexanoylglycine	28.59	0.08	0.27%	2.30	0.05	2.34%	1632	1648**
$Tartaric acid	29.34	0.08	0.26%	1.17	0.03	2.95%	1668	1665
3-Ureidopropionic acid	29.53	0.11	0.36%	2.95	0.06	2.13%	1663	1676
Xilose	30.50	0.07	0.24%	0.91	0.03	3.56%	1684	1645
$Xilitol	31.31	0.07	0.23%	0.89	0.01	1.57%	1695	1694
$Ribitol	31.75	0.07	0.21%	0.98	0.04	3.66%	1717	1712
cis-Aconitic acid	31.97	0.07	0.23%	1.44	0.04	2.45%	1751	1749
α-Glycerophosphate	32.66	0.07	0.20%	1.41	0.03	2.33%	1763	1751
Cl-Phenylalanine (ISTD 1)	33.03	0.08	0.23%	1.49	0.02	1.46%	1789	-
$Hippuric acid	33.93	0.08	0.23%	2.52	0.04	1.40%	1823	1820
$Citric acid	34.12	0.08	0.23%	1.31	0.03	2.61%	1840	1838
N-Phenyl acetyl glycine	35.03	0.08	0.23%	2.58	0.05	2.03%	1855	1848
$Fructose (1)	35.67	0.08	0.22%	0.95	0.03	3.40%	1867	1867
$Fructose (2)	35.95	0.08	0.23%	0.94	0.03	3.12%	1876	1875
$Glucose (1)	36.35	0.14	0.37%	0.99	0.02	2.18%	1891	1902
$Galactose (1)	36.70	0.10	0.26%	0.99	0.05	5.34%	1902	1883
$L-Tyrosine (2TMS)	36.71	0.06	0.16%	1.39	0.00	0.00%	1905	1934
$Galactose (2)	36.77	0.43	1.18%	0.93	0.04	4.81%	1918	1926
$Mannitol	37.22	0.08	0.23%	0.95	0.03	3.37%	1927	1925
p,p'-Dibromobiphenyl (ISTD 2)	37.80	0.08	0.21%	1.98	0.03	1.72%	1959	-
D-Gluconic acid	39.12	0.06	0.15%	1.03	0.01	1.30%	2037	1997
N-Acetyl-D-glucosamine	40.57	0.10	0.24%	1.50	0.04	2.55%	2121	2072
$Myo-inositol	40.69	0.06	0.14%	1.05	0.03	2.58%	2128	2126

$Uric acid	40.91	0.06	0.15%	1.36	0.04	3.06%	2130	2130
$L-Tryptophan	42.24	0.07	0.16%	1.59	0.03	1.89%	2181	2193
Stearic acid	43.27	0.06	0.14%	1.23	0.02	1.62%	2240	2236
Pseudouridine	46.07	0.06	0.14%	1.32	0.04	2.73%	2378	2375
Sucrose	52.23	0.09	0.16%	1.13	0.03	2.85%	2655	2653

$ analytes confirmed by reference compounds and external calibration
** reference I^T_S on standard apolar column (DB-1 and equivalents)